

Cognitive Diagnostic Research on Chinese Students' English Listening Skills and Implications on Skill Training

Huilin Chen¹ & Jinsong Chen²

¹ School of Education, Shanghai International Studies University, Shanghai, China

² Department of Psychology, Sun Yat-Sen University, Guangzhou, China

Correspondence: Huilin Chen, School of Education, Shanghai International Studies University, Shanghai, China.
E-mail: chlmailbox@163.com

Received: September 24, 2017 Accepted: November 6, 2017 Online Published: November 8, 2017

doi: 10.5539/elt.v10n12p107

URL: <http://doi.org/10.5539/elt.v10n12p107>

Abstract

By analyzing the test data of 2718 secondary school students in Guangzhou China on 15 listening items from Guangzhou English Achievement Examination (2015) through G-DINA model, the study explored the relationships among the listening comprehension skills. Based on the test specifications and listening skill taxonomies in existence, 5 experts in language skills and language testing conducted item content analysis independently for the 15 listening items, defined 5 listening attributes, and constructed the Q-matrix. After analyzing latent classes and their posterior probabilities, the study discovered the relationship among the listening skills. According to the listening skill relationship, the study provides insights on the sequence of listening skill training. The efficiency of training may be improved when closely related listening skills are instructed and practiced at the same time. The study also demonstrates that the compensatory and saturated G-DINA model caters to the characteristics of listening comprehension skills and can be applied to tests involving highly interactive and hierarchical skills.

Keywords: cognitive diagnosis, English achievement test, listening skill, G-DINA model, skill training

1. Introduction to Cognitive Diagnostic Research

Cognitive diagnostic assessment (CDA) is designed to measure specific knowledge structures and processing skills in students so as to provide information about their cognitive strengths and weaknesses (Leighton & Gierl, 2007). Cognitive diagnosis models (CDMs) are latent variable models developed primarily for cognitive diagnostic assessments to assess student mastery and non-mastery on a set of finer-grained skills and are developed to provide more targeted information in the form of score profiles that can allow for effective measurement of student learning and progress, designing of better instruction, and possibly intervention to address individual and group needs (de la Torre, 2009, 2011).

The main purpose of CDAs is to classify learners into unique attribute mastery profiles by calibrating tests with CDMs. CDMs not only are designed for CDAs but also can be applied to extracting diagnostic information from existing tests, if one can identify a set of fine-grained attributes that are useful for providing learners with diagnostic feedback. CDMs may differ in terms of model saturation, interattribute relationships, estimation methods, estimation software, and its versatility in dealing with polytomously scored items. These differences can actually have significant impact on the estimation of examinee skill mastery status and their interpretation (Lee & Sawaki, 2009). Model saturation determines whether a CDM allows for all possible item parameters, including interactions of attributes. A saturated CDM can not only include all single-skill attributes required by items but also take all possible attribute interactions as mixed-skill attributes. A reduced CDM only allows for item parameters of single-skill attributes. Interattribute relationships determine whether the probability of success in one attribute can influence that in other attributes required by the same item. Under a noncompensatory CDM, an item can be successfully answered only if all the required attributes for the item have been successfully mastered and executed. That is to say, one attribute cannot be completely compensated for by other attributes in terms of item performance. In contrast, under a compensatory CDM, successfully executing only a few or some of the attributes required for an item may achieve the correct response to that item. In other words, the attribute structure is compensatory in that strength in one attribute may compensate for weakness in another, thus mastery of all attributes involved in an item is not necessarily required for a test taker to answer the

item correctly.

Since Tatsuoka (1983) developed the first CDM which could estimate examinees' mastery levels of attributes, more than 60 CDMs of various formulations have been proposed in the psychometric literature. Examples of CDMs with wide recognition include the Rule Space Methodology (RSM; Tatsuoka, 1983), the deterministic inputs, noisy "and" gate (DINA; de la Torre, 2009; Junker & Sijtsma, 2001) model, the fusion model (Hartz, Roussos, & Stout 2002), the general diagnostic model (GDM; von Davier, 2005), and the generalized DINA (G-DINA; de la Torre, 2011) model. The development and applications of those CDMs have always been accompanied with tests on mathematics, medical science, and psychology.

2. Applications of CDMs to Language Tests

Encouraged by the success in applications of CDMs to mathematical, medical, and psychiatric tests, researchers begin to have interest in applying CDMs to language tests. Sheehen, Tatsuoka, & Lewis (1993) made an ETS report on applying RSM to analyzing the document processing skills of American adolescents. Buck, Tatsuoka, and Kostin (1997) applied RSM to analyzing the cognitive attributes in TOEIC reading items. Buck & Tatsuoka (1998) used the same model again to analyze the cognitive attributes in an open-ended English listening test. Unlike RSM adopted in the above studies, von Davier (2008) applied GDM model to analyzing the cognitive attributes in TOEFL reading and listening items; Jang (2009) applied fusion model to analyzing the cognitive attributes in the reading items of LanguEdge, a simulated TOEFL test; Lee & Sawaki (2009) made a more comprehensive study by applying GDM, the fusion model, and the latent class model respectively to analyzing the cognitive attributes in TOEFL reading and listening items. The study revealed that the three models produced similar results in terms of examinee classification, but some subtle differences between the results of GDM and those of other two models were identified as well.

Cognitive diagnosis of language tests is an important challenge in cognitive diagnosis research, which is determined by the characteristics of language skills and language tests. On the one hand, language tests are multidimensional, and most of the language tests are integrative tests, and the skills of integrative language tests are often multi-dimensional and hierarchical (Heaton, 1991); On the other hand, since language skills are more abstract and different language skills are linked to one another, language skills are more difficult to define and distinguish (Oller, 1979; Oller, & Kahn, 1981). Although there are some cognitive diagnosis research on language testing, most of the CDMs applied are reduced or non-compensatory models. The early RSM is a method of classification rather than a psychometric model as there is no item or person parameter to estimate. The GDM, the fusion model, and the latent class model are only reduced CDMs and the validation with fit measures was generally limited. Therefore, the previous cognitive diagnosis research on language testing may not cater to the characteristics of language skills and language tests and thus the diagnosis information retrieved from those studies may lack accuracy.

3. The G-DINA Model

The G-DINA model developed by Jimmy de la Torre (2011) relaxes the DINA model assumption of equal probability of success for all attribute vectors and is a saturated model. Without any constraints, the G-DINA model has $2^{K_j^*}$ parameters for item j , thus affording it greater generality compared to the DINA model whenever $K_j^* > 1$. Furthermore, the G-DINA model allows examinees with fewer required attributes for an item to achieve a certain probability of answering the item correctly so that the G-DINA model belongs to the compensatory CDM. The function of the G-DINA model based on $P(\alpha_{ij}^*)$ is as follows.

$$P(\alpha_{ij}^*) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk} \alpha_{lk} + \sum_{k'=k+1}^{K_j^*} \sum_{k=1}^{K_j^*-1} \delta_{jkk'} \alpha_{lk} \alpha_{lk'} \dots + \delta_{j12\dots K_j^*} \prod_{k=1}^{K_j^*} \alpha_{lk}$$

The function above can be decomposed into the sum of the effects due to the presence of specific attributes and their interactions. δ_0 represents the baseline probability (i.e., probability of a correct response when none of the required attributes is present), which can be regarded as the guessing parameter; δ_k is the change in the probability of a correct response as a result of mastering a single-skill attribute (i.e., α_k); $\delta_{kk'}$, a first-order interaction effect, is the change in the probability of a correct response due to the mastery of both α_k and $\alpha_{k'}$ that is over and above the additive impact of the mastery of the same two attributes; and $\delta_{12\dots K_j^*}$ represents the change in the probability of a correct response due to the mastery of all the required attributes that is over and above the

additive impact of the main and lower-order interaction effects (de la Torre, 2011). Since the G-DINA model is both compensatory and saturated, it may cater to the integrative and hierarchical features of language skills and language tests. Therefore, the G-DINA model was adopted as the CDM to analyze language skill structures in this study.

4. Research Design

This research took the listening subtest of Guangzhou English Achievement Examination as a case study. This study analyzed the 15 listening items from Guangzhou English Achievement Examination (2015). The items are all dichotomously scored items and related with 5 English conversations. The sample examinees in this study include 2718 secondary school students in Guangzhou China. Both the size of the sample and the number of items satisfy the requirement of this cognitive diagnostic analysis.

A cognitive diagnostic analysis usually starts with the identification of a set of attributes assessed in a test and the specification of the relationships between the attributes and test items. An attribute “refers to anything that affects performance on a task: either a task characteristic, or any of the knowledge, skills or abilities necessary to complete the task.” (Buck & Tatsouka, 1998: 121) A fundamental assumption in cognitive diagnosis is that each item on a given test can be described in terms of a set of attributes that should be mastered by an examinee to answer each item correctly (Gierl et al., 2000). The soundness of the attribute definition and item coding are the critical factors that determine the interpretability of attribute mastery profiles to be obtained from data analysis. There are mainly four sources which can be utilized to define attributes: test specifications, existing skill taxonomies, analysis of item content, and think-aloud protocol analysis of examinees’ test taking process. Once the attributes are defined for a particular test, a Q-matrix for that test can be constructed. The Q-matrix defines which attributes are assumed to be involved in answering each item correctly.

In this study, the Q-matrix is constructed through substantive analysis of item content which was recognized by Douglas, de la Torre, Chang, Henson, and Templin (2006). In the item content analysis of this study, 5 experts in language skills and language testing inspected the 15 listening items and independently coded each test item for the attribute(s) required to answer the given item correctly. The coding of attribute(s) for each item was also supplemented with references to the specifications of the test and existing listening skill taxonomies.

The listening skills defined in the specifications of Guangzhou English Achievement Examination are as follows:

- Guessing the meaning of words / phrases from context;
- Understanding the main idea and purpose;
- Obtaining specific information;
- Understanding the speaker's intentions, opinions and attitudes;
- Making inference;
- Recognizing discourse markers (Guangzhou Institute of Educational Research, 2011)

The existing listening skill taxonomies consulted in this study include Richard’s (1983) listening micro-skill taxonomy, Zou’s (2011) listening comprehension skill taxonomy, and Buck’s (2001) 3-skill default listening construct.

Based on the above listening specifications and taxonomies, the five experts conducted initial coding in a collective way by selecting the salient coding options for each item. After that, the experts were asked to discuss until they reached an agreement of five attributes for the whole listening subtest as shown in Table 1. Based on the five attributes, each expert conducted the second round of coding and constructed their own Q-matrices individually. For each item, the attributes selected by the majority of the experts were taken as the attributes required by that item. Finally, we came up with a common Q-matrix as shown in Table 2.

Table 1. Attribute definition

Listening attributes	Definition
Retrieving explicit information	Retrieving information the same as or similar to the correct answer to the listening items
Judging speaking situation	Judging the speaking situation and speaker characteristics
Generalizing multiple pieces of information	Generalizing multiple pieces of information scattered in the listening material and reaching a comprehensive understanding
Interpreting and transcribing explicit information	Interpreting and transcribing the explicit information in the listening material, understanding the concept and logical relationship embodied in the information.
Making inference	Understanding the information not explicitly stated by making inference or prediction

Some of the attributes defined by experts are equivalents of the skills defined in the test specifications. “Making inference” has the same counterpart in the test specifications, “Retrieving explicit information” is just “Obtaining specific information”, and “Generalizing multiple pieces of information” is similar to “Understanding the main idea and purpose”. Although “Judging speaking situation” and “Interpreting and transcribing explicit information” do not have equivalents in the test specifications, they can be implied in “Guessing the meaning of words / phrases from context” and “Understanding the speaker's intentions, opinions and attitudes” respectively. The Q-matrix coded by the 5 language experts is listed in Table 2.

Table 2. Q-matrix

	Retrieving explicit information	Judging speaking situation	Generalizing multiple pieces of information	Interpreting and transcribing explicit information	Making inference
1	1	1	0	0	0
2	1	0	1	1	0
3	1	0	0	1	1
4	0	1	1	0	0
5	1	0	0	1	0
6	1	0	0	0	0
7	1	0	1	1	0
8	1	0	1	0	0
9	0	0	1	0	1
10	0	1	1	0	0
11	1	0	0	1	0
12	1	0	0	1	0
13	1	0	0	1	0
14	1	1	0	0	0
15	1	1	0	0	0
Total	12	5	6	7	2

According to the Q-matrix above, the low level listening skill (Retrieving explicit information) accounts for a large proportion while the advanced listening skills (Making inference) plays a minor role in the test. The proportions the listening skills account for roughly match the actual situation how secondary school students master listening skills.

5. Results and Discussion

Based on the Q-matrix above, the test data of the 2718 examinees on the 15 listening items are analyzed with the G-DINA model code (Jimmy de la Torre, 2011) operated under OxEdit software (Doornik 2009). The absolute model fit of the analysis is based on the residual between the observed and predicted correlation of item pair with the Fisher transformation (ρ) and the residual between the observed and predicted log-odds ratios (LOR) of pair-wise item responses (l) jointly (Chen, de la Torre, & Zhang 2012). At a certain significant level, if the Max z-Scores based on ρ and l are larger than the critical values (CV) based on ρ and l respectively, the CDM adopted in the analysis will be rejected. The higher the significant level, the fitter the CDM.

The absolute model fit statistics for this study is shown in Table 3.

Table 3. Absolute model fit

		Prop	ρ	l
Max Z Score		1.6783	3.2278	3.1464
	$P=$.01	3.4029	3.9024	3.9024
Zc Score (Bonferroni Correction):	$P=$.05	2.9352	3.4938	3.4938
	$P=$.10	2.7131	3.3042	3.3042

Table 3 shows that when the Q-matrix is adopted, the absolute model fit under G-DINA model can reach a high significant level at .10, which demonstrates that the Q-matrix defined by experts and the G-DINA model can be adopted for the data analysis.

Since the absolute model fit for the Q-matrix defined by experts and the G-DINA model reaches the significant level, further cognitive diagnostic analysis can be carried out.

According to the analysis of attribute prevalence, the subjects' mastery probability of each attribute can be obtained. Table 4 shows the results of attribute prevalence for the 2718 subjects.

Table 4. Attribute prevalence

Cognitive attribute	Mastery probability	Times of measurement
Retrieving explicit information	.66	12
Judging speaking situation	.54	5
Generalizing multiple pieces of information	.66	6
Interpreting and transcribing explicit information	.37	7
Making inference	.57	2

According to the table, "Retrieving explicit information" and "Generalizing multiple pieces of information" are mastered best by the subjects, and "Interpreting and transcribing explicit information" is most poorly mastered by the subjects.

According to the analysis of latent classification, the subjects' mastery types of attributes (latent classes) and their posterior probabilities can be obtained. Table 5 shows the 17 mastery types of attributes whose posterior probabilities are higher than 1% of the total sum of all posterior probabilities.

Table 5. Latent classes

Latent class	Posterior probability
"11111"	0.3411
"10001"	0.1279
"00110"	0.0928
"00100"	0.0928
"10000"	0.0456
"01010"	0.043
"01000"	0.043
"10101"	0.0341
"11101"	0.0327
"11001"	0.0262
"00000"	0.0172
"00010"	0.0172
"11100"	0.0153
"01001"	0.0153
"01011"	0.0153
"11110"	0.0132
"10100"	0.012
Sum	0.9847

The 5 figures as a whole representing the latent class symbolize “Retrieving explicit information”, “Judging speaking situation”, “Generalizing multiple pieces of information”, “Interpreting and transcribing explicit information”, and “Making inference” from the left to the right respectively. Four dominant mastery types of attributes whose posterior probabilities are higher than 9% can be discovered in Table 5. They are "11111", "10001", "00110", and "00100" in a descending order.

According to latent classes and posterior probabilities shown in Table 5, the 4 dominant latent classes of listening comprehension attributes demonstrate that there exist 4 dominant structures of listening comprehension skills in the cognition of the subjects. All the 5 attributes can be found in the 4 dominant structures, which demonstrates that the 5 attributes are representative components of the listening skill structure of the subjects.

By analyzing the occurrence of the attributes in the 4 dominant latent classes, each of which accounts for more than 9% of the sum of posterior probabilities of all latent classes, the interrelationships among the listening comprehension skills can be easily revealed. The largest latent class contains all of the 5 attributes, which demonstrates the fact that the 5 attributes are closely interrelated. Since the “11111” latent class contains all attributes which are interrelated with one another, the “11111” latent class has the largest posterior probability. The "00100" latent class is the only single attribute latent class among the dominant latent classes, which demonstrates “Generalizing multiple pieces of information” is the most independent skill among the 5 listening skills and can be mastered almost singly.

The structure of the listening comprehension skills can be refined further by taking other latent classes into consideration. It can be found that single attribute latent classes "10000" and "01000" both have posterior probabilities over .04, which demonstrates that “Retrieving explicit information” and “Judging speaking situation” are to some extent independent but may still have relationships with other skills. Another single attribute latent classes which has a posterior probability over .01 is "00010", which demonstrates that “Interpreting and transcribing explicit information” is highly dependent and has strong relationship with other skills. The only single attribute latent class which has a posterior probability below .01 is “00001”, which demonstrates that “Making inference” is the most dependent skill and can only be mastered together with other skills. The latent class "00000" has a posterior probability below .02, which demonstrates that almost all the subjects have fairly good mastery of listening skills involved in the test. The skill having closest relationship

with “Making inference” is “Retrieving explicit information” because of the latent class "10001" whose posterior probability is over .12. The skill having closest relationship with “Interpreting and transcribing explicit information” is “Generalizing multiple pieces of information” because of the latent class "00110" whose posterior probability is over .09. The skill having closest relationship with “Judging speaking situation” is “Interpreting and transcribing explicit information” because of the latent class "01010" whose posterior probability is over .04.

6. Implications on Listening Skill Training

The structure of the relationship among listening comprehension skills can provide some insights on the arrangement of listening skill training. Since the most dependent skill “Making inference” has closest relationship with “Retrieving explicit information”, the training of “Retrieving explicit information” can be regarded as the prerequisite to the training of “Making inference” simply because only after detecting the explicit language forms and understanding superficial meanings of those forms in the listening process can students make inference about implicit information. Since the highly dependent skill “Interpreting and transcribing explicit information” has closest relationship with “Generalizing multiple pieces of information”, the training of “Generalizing multiple pieces of information” can be regarded as the prerequisite to the training of “Interpreting and transcribing explicit information” probably because “Generalizing multiple pieces of information” prepares for interpretation. Since the highly dependent skill “Interpreting and transcribing explicit information” also has close relationship with the somewhat independent skill “Judging speaking situation”, the training of “Judging speaking situation” should also be conducted before the training of “Interpreting and transcribing explicit information” probably because “Judging speaking situation” is a simple form of interpretation. Furthermore, the posterior probabilities of "10001", "10101", and "11101" are ranked from high to low, which demonstrates that the training of “Generalizing multiple pieces of information” should be conducted after the mastery of both “Retrieving explicit information” and “Making inference”, then the training of “Judging speaking situation”, and finally the training of “Interpreting and transcribing explicit information”. Therefore, the order of listening skill training can be expressed in Figure 1.

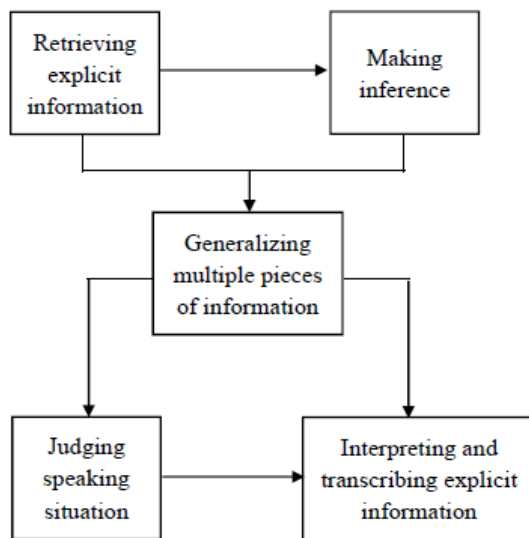


Figure 1. Order of listening skill training

7. Conclusion

This study adopted G-DINA model to explore the relationships among listening comprehension skills. This study analyzed the test data of 2718 secondary school students in Guangzhou China on 15 listening items from Guangzhou English Achievement Examination (2015) and explored the relationships among the 5 listening comprehension skills defined by experts. By analyzing latent classes and their posterior probabilities, the study discovered the relationship among the listening skills. The findings provide some insights on the sequence of listening skill training for the sample examinees. The closely related listening skills can be instructed and practiced at the same time to achieve efficiency, which may suggest the ideal order of listening skill training.

Furthermore, the study also demonstrates that the compensatory and saturated G-DINA model caters to the characteristics of listening comprehension skills and can be applied to tests involving highly interactive and hierarchical skills.

Acknowledgements

This paper is a partial fulfilment of the project ‘A Cognitive Study of the English Skill Structure of Chinese English Learners’ (17BYY101) sponsored by National Social Science Foundation of China, and is also supported by “TEM Examination Skill Bank Construction” and “Foreign Language Skill Assessment Reform” under Shanghai Foreign Language and Literature Discipline Construction Project.

References

- Buck, G. (2001). *Assessing Listening*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511732959>
- Buck, G., & Tatsuoka, K. (1998). Application of the rule-space procedure to language testing: Examining attributes of a free response listening test. *Language Testing*, 15, 119-157. <https://doi.org/10.1191/026553298667688289>
- Buck, G., Tatsuoka, K., & Kostin, I. (1997). The subskills of reading: Rule-space analysis of a multiple-choice test of second language reading comprehension. *Language Learning*, 47, 423-466. <https://doi.org/10.1111/0023-8333.00016>
- Chen, J., de la Torre, J., & Zhang, Z. (2013). Relative and absolute fit evaluation in cognitive diagnosis modeling. *Journal of Educational Measurement*, 50, 123-140. <https://doi.org/10.1111/j.1745-3984.2012.00185.x>
- de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, 34, 115-130. <https://doi.org/10.3102/1076998607309474>
- de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76, 179-199. <https://doi.org/10.1007/s11336-011-9207-7>
- Doornik, J. A. (2009). *An object-oriented matrix language Ox 6*. London: Timberlake Consultants Press.
- Douglas, J., de la Torre, J., Chang, H., Henson, R., & Templin, J. (2006). *Skills diagnosis with latent variable models*. Symposium presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.
- Gierl, M., Leighton, J. P., & Hunker, S. M. (2000). Exploring the logic of Tatsuoka's rule-space model for test development and analysis. *Educational Measurement: Issues and Practices*, 19, 34-44. <https://doi.org/10.1111/j.1745-3992.2000.tb00036.x>
- Guangzhou Institute of Educational Research. (2011). *A Guide for Guangzhou Junior High School Achievement Examination: English*. Guangzhou: Guangdong Education Press.
- Hartz, S. (2002). *A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality (Unpublished doctoral dissertation)*. University of Illinois at Urbana-Champaign.
- Heaton, J. B. (1991). *Writing English language test*. New York: Longman.
- Jang, E. E. (2009). Cognitive diagnostic assessment of L2 reading comprehension ability: Validity arguments for fusion model application to LanguEdge assessment. *Language Testing*, 26, 031-073. <https://doi.org/10.1177/0265532208097336>
- Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25, 258-272. <https://doi.org/10.1177/01466210122032064>
- Lee, Y., & Sawaki, Y. (2009). Application of three cognitive diagnosis models to ESL reading and listening assessments. *Language Assessment Quarterly*, 6, 239-263. <https://doi.org/10.1080/15434300903079562>
- Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2004). The attribute hierarchy model for cognitive assessment: A variation on Tatsuoka's rule-space approach. *Journal of Educational Measurement*, 41, 205-237. <https://doi.org/10.1111/j.1745-3984.2004.tb01163.x>
- Oller, J. W. (1979). *Language Tests at School: A Pragmatic Approach*. London: Longman.
- Oller, J. W., Jr. & Kahn, F. (1981). Is there a global factor of language proficiency? In J. A. S. Read (Ed.), *Directions in Language Testing*. (pp. 1-25). Singapore: Singapore University Press.

- Richards, J. C. (1983). Listening comprehension: approach, design, procedure. *TESOL Quarterly*, 17(2), 219-240. <https://doi.org/10.2307/3586651>
- Sheehan, K. M., Tatsuoka, K. K., & Lewis, C. (1993). *A diagnostic classification model for document processing skills (ETS Research Rep. No. RR-93-99)*. Princeton, NJ: Educational Testing Service. <https://doi.org/10.21236/ADA273790>
- Tatsuoka, K. (1983). Rule space: an approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20, 345-354. <https://doi.org/10.1111/j.1745-3984.1983.tb00212.x>
- von Davier, M. (2005). *A General Diagnostic Model Applied to Language Testing Data* (ETS Research Report RR-05-16). Princeton: Educational Testing Service.
- von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, 61, 287-307. <https://doi.org/10.1348/000711007X193957>
- Zou, S. (2011). *An Introduction to English Language Testing*. Beijing: Higher Education Press.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).